Title
Constructing kinetic models to elucidate structural dynamics of a complete RNA polymerase II elongation cycle

Journal
Physical Biology, 12(1)

ISSN
1478-3967

Authors
Yu, Jin
Da, Lin-Tai
Huang, Xuhui

Publication Date
2015

DOI
10.1088/1478-3975/12/1/016004

Peer reviewed

Constructing kinetic models to elucidate structural dynamics of a complete RNA polymerase II elongation cycle

2015 Phys. Biol. 12 016004

(http://iopscience.iop.org/1478-3975/12/1/016004)

Physical Biology

OPEN ACCESS

RECEIVED
7 September 2014

ACCEPTED FOR PUBLICATION
7 November 2014

PUBLISHED
05 December 2014

PAPER

Constructing kinetic models to elucidate structural dynamics of a complete RNA polymerase II elongation cycle

Jin Yu[1], **Lin-Tai Da**[2,3] **and Xuhui Huang**[3]

[1] Beijing Computational Science Research Center, Beijing, 100084, People's Republic of China
[2] Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA
[3] Department of Chemistry, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, People's Republic of China

E-mail: jinyu@csrc.ac.cn **and** xuhuihuang@ust.hk

Keywords: RNA polymerase II, kinetic modeling, single molecule experiment, molecular dynamics simulation, Brownian ratchet

Supplementary material for this article is available online

Abstract

The RNA polymerase II elongation is central in eukaryotic transcription. Although multiple intermediates of the elongation complex have been identified, the dynamical mechanisms remain elusive or controversial. Here we build a structure-based kinetic model of a full elongation cycle of polymerase II, taking into account transition rates and conformational changes characterized from both single molecule experimental studies and computational simulations at atomistic scale. Our model suggests a force-dependent slow transition detected in the single molecule experiments corresponds to an essential conformational change of a trigger loop (TL) opening prior to the polymerase translocation. The analyses on mutant study of E1103G and on potential sequence effects of the translocation substantiate this proposal. Our model also investigates another slow transition detected in the transcription elongation cycle which is independent of mechanical force. If this force-independent slow transition happens as the TL gradually closes upon NTP binding, the analyses indicate that the binding affinity of NTP to the polymerase has to be sufficiently high. Otherwise, one infers that the slow transition happens pre-catalytically but after the TL closing. Accordingly, accurate determination of intrinsic properties of NTP binding is demanded for an improved characterization of the polymerase elongation. Overall, the study provides a working model of the polymerase II elongation under a generic Brownian ratchet mechanism, with most essential structural transition and functional kinetics elucidated.

1. Introduction

RNA polymerase II (Pol II) is the core enzyme that catalyzes gene transcription in eukaryotic cells [1–4]. The multi-subunit Pol II consists of 12 subunits and is weighted about 550 kD. It works with various transcription factors to control multiple stages of transcription [5–10], from initiation, elongation to termination. High-resolution structures of the Pol II complex have been captured at different conformational states during its elongation cycle [11–20]. In particular, two prominent structural elements close to the active center of Pol II are well characterized as keys to the elongation control: one is a trigger loop (TL) that opens after product release and during polymerase translocation, and closes upon nucleotide binding and through catalysis [15, 21–26]; the other is a bridge helix (BH) that locates next to TL and assists translocation or active site re-arrangement [27, 28]. Pol II is one of the most studied multi-subunit RNA polymerases. Both ensemble and single molecule measurements had been conducted to investigate the elongation properties of Pol II [29–36]. Nevertheless, detailed mechanisms of Pol II elongation are still lack of or remain controversial.

The multi-subunit RNA polymerases have been suggested to work as the Brownian ratchet (BR) along the DNA track, as the translocation takes place spontaneously back and forth between the pre- and post-translocation states [37–39]. The NTP binding to the active site prevents backward movements of the RNA polymerase, and therefore promotes forward

movements. Remarkably, the base-pair stepping motion of a similar multi-subunit RNA polymerase from *E. coli* was resolved at the single molecule level [37]. The experimental fittings to the force–velocity data of the single enzyme elongation supported a BR model, though NTP binding was suggested to happen at either pre- or post-translocation.

Similar features have been derived from a recent experiment using an optical-trapping assay with high spatiotemporal resolution to probe single yeast Pol II elongation [34]. Global fits to the force–velocity data again supported the above BR model with NTP binding at either pre- or post-translocation. The model is later on referred as a branched BR model. It should be noted that the translocation or the force-dependent transition is assumed to happen much faster than subsequent kinetic steps in each elongation cycle in this branched BR model [34].

In a most recent single molecule experiment challenging the individual yeast Pol II with nucleosomal barrier, however, the above assumption of fast translocation turns out to be unnecessary [35]. The study sought to achieve comprehensive kinetic characterization of Pol II elongation without making any assumption about the rate-limiting step of the elongation [35]. In the experiment, an optical tweezers assay was used to follow a single Pol II elongation at varying NTP conditions, and under an assisting or opposing applied force, similarly as that in [34]. In addition, different tracks of DNA, bare or with nucleosomal barrier, were both examined for the Pol II elongation [35]. Based on similar force measurements as that in [34] but an alternative way of data fitting, it has been shown that the forward translocation rate can be very low (∼88 s^{-1}) and comparable to the force-independent slow rate or catalytic rate (∼35 s^{-1}) [35]. The finding suggested that a non-branched generic BR model holds as long as the translocation, or indeed, the force-dependent transition prior to NTP binding, is also regarded as a slow step during the transcription elongation.

On the computational side, molecular modeling and simulations based on high-resolution structures of Pol II have provided fruitful insights into atomic level mechanisms of transcription elongation in recent years [23, 26, 40–48]. For example, normal mode analysis employing atomistic force field to the polymerase structure suggested that productive translocation requires TL opening, while the translocation is further inhibited by the presence of an NTP in the active site [41]. At the same time, molecular dynamics (MD) simulation on atomistic structures of Pol II also suggested that the thermally driven translocation of the polymerase is only achievable with an open TL [42].

Furthermore, extensive all-atom MD simulations employing the Markov state model (MSM) techniques have recently captured fast dynamics of pyrophosphate ion (PPi) release at microsecond time scale [43, 44]. PPi release happens after chemical addition of an inserted NTP; after the PPi release translocation of the polymerase happens. The study suggested that the PPi release could induce a slight tip opening motion of TL, even through a full TL opening is yet to happen. Remarkably, the atomistic MD simulations employing the same techniques while focusing on translocation of Pol II reported that translocation in the absence of NTP can occur at tens of microseconds [48]. The results appear inconsistent with the slow translocation step (about ten milliseconds) reported by the single molecule experiments [35]. However, it is noted that the pre-translocation structure adopted in the MD simulation has already an open TL configuration.

Combing data and analyses at single molecule level from both the experimental and computational sides, we propose here a working model of Pol II elongation, in which fast translocation of the RNA polymerase is preceded by a slow process of TL opening. We regard that the *force-dependent* transition detected in the single molecule experiments [35] as a combination of the slow TL opening and the fast translocation. The model resolves seemingly controversies from the above studies. At the same time, we also show that there are at least two scenarios for the *force-independent* rate-limiting transition that follows the translocation and NTP binding in the elongation. In addition, we discuss how sequence stability variation along DNA affects the elongation rate and explain why that cannot induce the detected force-dependent slowing down. We also estimate intrinsic NTP dissociation constants of the polymerase so that to infer which scenario is more likely for the force-independent transition.

2. Model constructs and results

2.1. The generic BR model and single molecule measurements supporting this non-branched model

First we address the elongation kinetics of Pol II in a minimal three-state representation: the Pol II follows a generic BR mechanism, and NTP binding serves as a pawl and is only allowed at *post*-translocation. In the depicted scheme in figure 1(a), state I, II, III are the pre-translocated or pre-trans, post-translocated or post-trans, and NTP-bound or substrate state. Translocation happens in between state I and II, with both forward and backward transition rates depending on the mechanical force for or against the movements [34, 35]. Basically, transition I → II encloses a force-dependent process that happens after the product formation but prior to the NTP binding. The rate of NTP binding (II → III) is proportional to the NTP concentration, as the binding process is presumably dominated by NTP diffusion in the solution. On the other hand, the product formation step (III → I) lumps several events together: a potentially slow process of waiting (such as the active site rearranging or tightening [42]), followed by chemical addition of

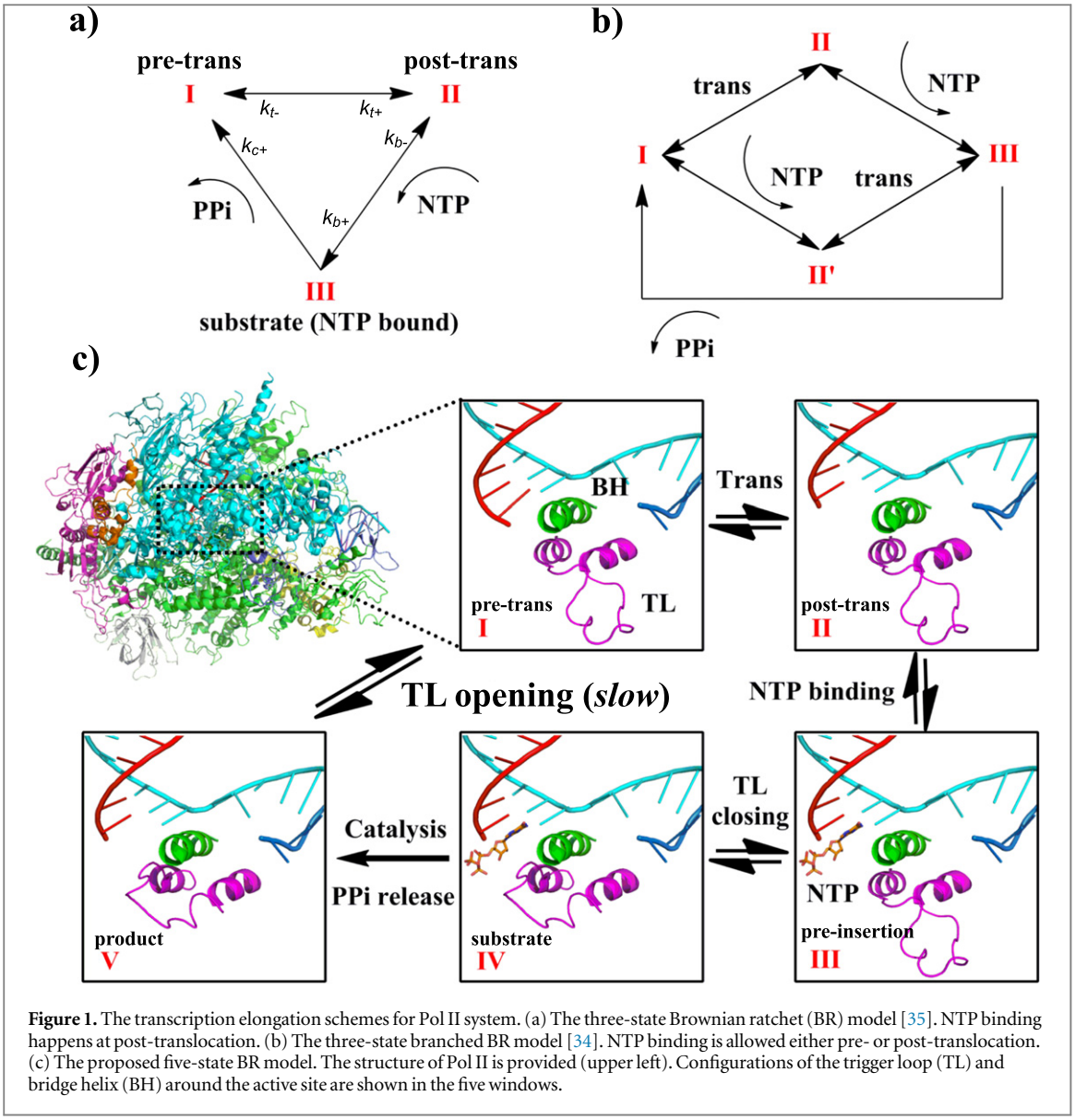


Figure 1. The transcription elongation schemes for Pol II system. (a) The three-state Brownian ratchet (BR) model [35]. NTP binding happens at post-translocation. (b) The three-state branched BR model [34]. NTP binding is allowed either pre- or post-translocation. (c) The proposed five-state BR model. The structure of Pol II is provided (upper left). Configurations of the trigger loop (TL) and bridge helix (BH) around the active site are shown in the five windows.

the nucleotide, and then the PPi release. When only one of these events is significantly slow, it is a good approximation to model the full product formation process as one transition. In addition, if PPi concentration is sufficiently low, the above transition approaches irreversible.

When the off-pathway transition such as transcription pausing or backtracking is not considered, the elongation rate of the above scheme is derived as:

$$v = \frac{k_{\max}\,[\mathrm{NTP}]}{K\left[1+\sigma_t(F)\right]/\left(1+k_c^+/k_t^+\right)+[\mathrm{NTP}]}, \quad (1)$$

where $k_{\max} = k_c^+/(1 + k_c^+/k_t^+(F))$, with $k_t^+(F)$ and k_c^+ the forward rates of the *force-dependent* (translocation I → II) and the *force-independent* (catalytic III → I) step, respectively, in the three-state scheme (figure 1(a)). In particular, $K \equiv K_d + k_c^+/k_b^0$ ($K \sim K_d$ when k_c^+/k_b^0 is small), where K_d is the intrinsic NTP dissociation constant ($K_d = k_b^-/k_b^0$, with $k_b^+ = k_b^0\,[\mathrm{NTP}]$ and k_b^- the 'on' and 'off' rate of NTP), and $\sigma_t(F) = k_t^-/k_t^+ = \sigma_t^0 \exp[-F\Delta/k_B T]$ is the force-dependent bias against the forward translocation, where $\sigma_t^0 \equiv k_{t(F=0)}^-/k_{t(F=0)}^+$ is defined at zero force (Δ is 1 nt distance; $F > 0$ for an assisting force). $\sigma_t^0 \sim 1$ is expected for the Brownian motion. As a result, how fast the elongation rate approaches to saturation (measured by $K_M \equiv K\,[1 + \sigma_t(F)]/(1 + k_c^+/k_t^+(F))$), as well as how large the saturation rate ($k_{\max} = k_c^+/(1 + k_c^+/k_t^+(F))$) is, depends on the implemented force F in the single molecule manipulation.

In the single molecule studies by Dangkulwanich *et al* [35], no particular assumption (such as translocation being very fast) was made to the kinetic rates in the BR scheme (figure 1(a)), which was used to fit the experimental data. Among their results, it was measured $k_{\max} = k_c/(1 + k_c^+/k_t^+(F)) = 25 \pm 3$ nt s^{-1} and $K_M \equiv K\,[1 + \sigma_t(F)]/(1 + k_c^+/k_t^+) = 39 \pm 12$ μM, according to the Michaelis–Menten fitting to the pause-free elongation velocity versus [NTP] data (at an applied force of 6.5 pN) [35]. To determine the

respective values of k_t^+ and k_c^+, additional measurements challenging Pol II with nucleosome barriers were conducted [35]. By comparing the pause-free velocities on nucleosome DNA and bare DNA, it showed that both the force-dependent and force-independent rates are low ($k_{t(F=0)}^+ = 88 \pm 23\ s^{-1}$ at zero force, and $k_c^+ = 35 \pm 3\ s^{-1}$) [35]. Hence, the translocation step seems to be another slow step aside from the catalytic one. Furthermore, pause densities during elongation were measured at different NTP concentrations, giving $k_t^- \cdot K \sim (4.7 \pm 0.5) \times 10^3\ \mu M^{-1}\ s^{-1}$. As a result, it was estimated that for the backward translocation $k_{t(F=0)}^- \sim 680\ s^{-1}$ ($\sigma_t^0 \sim 7.7$), hence $K \sim 9.2\ \mu M$ as $K \equiv K_d + k_c^+/k_b^0$ [35].

2.2. A branched BR model and single molecule measurements supporting this branched model

For the pause-free elongation rate obtained in the generic BR scheme (figure 1(a)) and equation (1), one can see that by assuming $k_t^+ \gg k_c^+$, $k_{\max}$ is reduced to the catalytic rate k_c^+. That says, when the force-dependent transition happens much faster than the force-independent step in this scheme, the saturation rate of the elongation becomes that of the force-independent step. However, in single molecule experiments of Pol II [34, 35], force dependency of the saturation elongation rate had always been detected, in particular, in the case of the mutant polymerase E1103G [34]. Therefore, if one adopts the assumption $k_t^+ \gg k_c^+$ while using the non-branched BR scheme (figure 1(a)), the model could not fit well to the single molecule experimental data, as shown in [34].

In order to fit the experimental data under the fast translocation assumption, a branched BR model was adopted [34], in which NTP binding is allowed either at pre-translocation or at post-translocation (see figure 1(b)). As $k_t^+ \gg k_c^+$ holds, the pause free elongation rate of the branched BR model is

$$v = \frac{\dfrac{k_{\max}}{1 + \sigma_t(F)}[\mathrm{NTP}]}{\dfrac{K'}{1 + \sigma_t(F)} + [\mathrm{NTP}]} \qquad (2)$$

such that $k_{\max} = k_c^+/(1 + k_c^+/k_t^+(F)) \sim k_c^+$ under the above assumption, while $K' \equiv K_d[1 + \sigma_t(F)] + k_c^+/k_b^0$. Then the saturating elongation rate $\frac{k_c^+}{1+\sigma_t(F)}$ becomes force dependent under this branched scheme. Hence, it explains why the experimental data fitted better with the branched BR model than with the non-branched one, under the fast translocation assumption [34].

Correspondingly, three parameters were obtained for the wild-type Pol II from fitting to the branched BR scheme in [34]: $k_c^+ = 34 \pm 2\ s^{-1}$, $\frac{K'}{1+\sigma_t(F)} = 140 \pm 16\ \mu M$, and $\sigma_t^0 = 0.2 \pm 0.1$. The data fitting from [34], along with the same fitting to the data measured later [35], is shown in SI figure S1.

If one compares the fitted parameters from the branched BR model [34] with that from the non-branched BR model [35], one sees that except for an identical catalytic rate k_c^+, the other key features are obtained quite differently: (i) in the branched BR model, $\sigma_t^0 \sim 0.2$, so the post-translocation state is more populated than the pre-translocated state. In the non-branched BR model, however, $\sigma_t^0 \sim 7.7$, so it is the pre-translocated state (under the three-state scheme) that is more populated. (ii) In the branched BR model, the intrinsic NTP dissociation constant can be valued high as $K_d < K'/(1 + \sigma_t) \sim 140\ \mu M$; while the value of K_d is restricted to be quite low as $K_d < K = K_d + k_c/k_T^0 \sim 9.2\ \mu M$ in the non-branched three-state BR model. Hence, one sees that even though the pause-free elongation velocities were measured consistently on Pol II in both experiments [34, 35], different choices of the data-fitting models lead to quite different physical interpretations of the system.

Based on structural and dynamical properties of Pol II revealed from computational studies, we present an expanded elongation scheme (shown in figure 1(c)) in this work. The scheme follows the generic non-branched BR model, while five kinetic states instead of three are used to provide slightly more specific descriptions of the Pol II elongation. Essentially, we propose a force-dependent TL opening process prior to the translocation. We justify this proposal below.

2.3. TL opening is a necessary step prior to translocation and can be slow and force dependent

In contrast to the single molecule experiment that suggested a slow translocation step of Pol II at tens of milliseconds, recent atomistic MD simulations implementing the MSM techniques identified the Pol II translocation at tens of microseconds in the absence of incoming NTPs [48]. This study supported the BR model where Pol II can move between the pre- and post-translocation states with nearly identical transition rates. The translocation rates ($\sim 10^5\ s^{-1}$) [48], however, are much larger than the experimentally measured force-dependent rate ($\sim 10^2\ s^{-1}$) [35]. Though including a full transcription bubble into the simulation may result in a longer translocation time, it is worth pointing out that the experimentally measured force-dependent rate does not necessarily apply to the translocation per se. The single molecule experimental study on the mutant Pol II E1103G [35] (addressed below) strongly indicates a conformational change related to TL rather than the transcription bubble that leads to the slow and force-dependent transition in the generic BR scheme (figure 1(a)). If the conformational change happens as a pre-requisite right before the translocation, then an assisting force of the translocation can easily accelerate the conformational change (or inhibit the change in the reversed change direction); vice versa, an opposing force of the

translocation may hinder the preceding conformational change (or facilitate it in the reversed change).

In previous computational studies of Pol II, it had been suggested that TL opening is required for the Pol II translocation to happen [41, 42]. Using normal model analyses, it was found that the reduced flexibility of the clamp domain upon TL closing or NTP binding to the active site translates into reduced mobility of the downstream DNA, thereby, effectively inhibits the translocation [41]. Further, it was shown that MD simulations with open and closed TLs sampled a common state with slight forward translocation relative to the x-ray structures, but more significant forward translocation was only observed in simulations with an open TL [42].

Another important event that happens after catalysis and prior to translocation is PPi release. MD simulation studies have shown that dynamics of the PPi release is much faster than the complete opening motion of TL. Nevertheless, the PPi release can increase the flexibilities of the tip region of the TL domain [43, 44]. On the other hand, in simulating the Pol II translocation, the pre-translocated structure adopted at the beginning of the simulation already has an *open* TL (modeled from the crystal structures: PDB id: 1I6H, 2NVT and 2E2J) [48]. Hence, it seems that a closed TL puts a 'brake' on the downstream DNA until after the PPi release. The full opening of the TL after the PPi release removes the brake and allows the translocation to proceed. Under an assisting force, the TL opening would be accelerated as the downstream DNA is forced to move backward. That is, the TL opening can be regarded as a slow and force-dependent transition that allows the fast translocation to happen.

Essentially, the single molecule study on the mutant E1103G indicated that the rate of the slow force-dependent transition would drop significantly (from 88 s^{-1} to 44 s^{-1}) upon the mutation [35]. E1103 actually locates on one end of the TL. The finding strengthens the idea that conformational changes related to TL rather than the transcription bubble cause the force-dependent slowing down. In the TL closed structure, we notice that salt bridges form between the negatively charged E1103 and two positively charged residues, R1100 and K1112, with distances at ~4.4 Å and 3.5 Å, respectively (see figure 2(a)). The geometry of the salt bridge between E1103 and R1100, however, appears to be *suboptimal*, compared to regular salt bridges (<4 Å). The impact of salt bridge interactions on the helix folding and unfolding kinetic has been closely examined in a recent study using temperature-jump transient-infrared spectroscopy and steady-state UV circular dichroism [49]. In that study, the effect of Glu–Arg salt bridges on the kinetics of alpha-helix folding was investigated, which shows that suboptimal salt bridges with unfavorable geometry kinetically destabilize the folded structure or promote the helix unfolding. It is then likely that in our study of Pol II, the suboptimal salt bridge (E1103-R1100) formed in the closed configuration of TL destabilizes the closed/folded TL, or promotes the TL opening/unfolding. In contrast, the mutation (E1103G) that abolishes the destabilizing salt bridge (E1103-R1100) thus brings a relatively stabilized form of the closed TL, or reduces the TL opening rate comparing to the wild type.

Following the above analyses, we use a five-state BR model to describe the elongation kinetics of Pol II. The reasons to choose five states are: (i) upon the three basic steps (translocation, NTP binding, and catalysis), we want to separately model the TL opening process as a slow transition, after the catalytic product formation, but prior to the translocation. Hence, at least four kinetic steps/states are needed for an elongation cycle. (ii) In addition, we want to consider and compare two scenarios for the *force-independent* slow event: the event either happens right upon NTP binding (model A below), or happens at least one kinetic step further down the reaction path (model B). To that end, two kinetic steps are considered between the NTP binding/pre-insertion state and the product state as the simplest case. Hence, five states/steps are now modeled. (iii) Since PPi release happens very fast and is off our concern in this work, we do not separately model the PPi release aside from the catalysis transition. As such, a five-state kinetic model serves for a minimal representation for the purpose of current study.

2.4. Model A: the five-state BR model with a rate-limiting TL closing/ isomerization upon NTP binding

In the five-state scheme (figure 1(c)), state I–V refer to the pre-trans, post-trans, pre-insertion [50], substrate, and product states. Correspondingly, translocation proceeds through I → II, NTP binding II → III, TL closing (or isomerization) III → IV, catalysis IV → V (including pre-catalytic adjustment, phosphoryl transfer, and PPi release), and TL opening V → I that allows for a next cycle. We assume in current scheme that (a) IV → V approaches irreversible (with forward rate k_c^+, and backward rate $k_c^- \to 0$) as PPi concentration is very low; (b) both V ↔ I (TL opening and closing prior to translocation, at rates $k_{\mathrm{TLo}}^{\pm}$) and I ↔ II (translocation forward and back, at rates $k_t^{\pm}$) are force dependent, with V ↔ I sufficiently slow, and I ↔ II very fast; (c) in particular for model A, III → IV is set as the rate-limiting step (forward rate or the rate of the TL closing after NTP is k_{TLc}^+, without force dependence), such that the transition rate $k_{\mathrm{TLc}}^{\pm} \ll k_c^+$ (the catalytic rate).

Correspondingly, one can write down a master equation for the probability/population distribution of the five kinetic states, using the population vector $\Pi = (P_{\mathrm{I}}\, P_{\mathrm{II}} P_{\mathrm{III}} P_{\mathrm{IV}} P_{\mathrm{V}})^T$

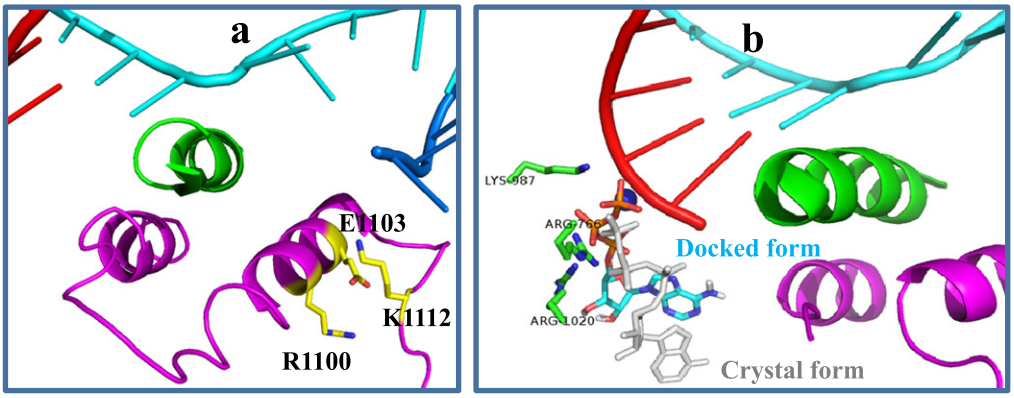


Figure 2. (a) Trigger loop structure at its closed/folded state. E1103 forms salt bridge interactions with R1100 and K1112, respectively. The salt bridge interactions can impact on the opening/unfolding kinetics of TL prior to the translocation. (b) Docking of the substrate ATP to the pre-translocation state (PDB id: 1I6H). The docked form of ATP is compared with its crystal form, which is bound with the post-translocation state of Pol II (PDB id: 1R9T). The Mg^{2+}A is shown in blue sphere, three positively charged residues: Lys987, Arg766 and Arg1020 in the active site are also highlighted. The docking was performed using Autodock4 [64].

$$\frac{\mathrm{d}}{\mathrm{d}t}\Pi = M\Pi, \tag{3}$$

where M stands for the transition rate matrix:

$$M = \begin{pmatrix} -k_t^+ - k_{\mathrm{TLo}}^- & k_t^- & 0 & 0 & k_{\mathrm{TLo}}^+ \\ k_t^+ & -k_t^- - k_b^+ & k_b^- & 0 & 0 \\ 0 & k_b^+ & -k_b^- - k_{\mathrm{TLc}}^+ & k_{\mathrm{TLc}}^- & 0 \\ 0 & 0 & k_{\mathrm{TLc}}^+ & -k_{\mathrm{TLc}}^- - k_c^+ & k_c^- \\ k_{\mathrm{TLo}}^- & 0 & 0 & k_c^+ & -k_c^- - k_{\mathrm{TLo}}^+ \end{pmatrix}.$$

In principle, there are $5\times5=25$ transitions allowed in a network representation. For a commonly used linear kinetic scheme, branched transitions are not considered. Solving equation (3) at steady state ($\frac{\mathrm{d}}{\mathrm{d}t}\Pi = M\Pi = 0$), under the model A assumptions (see SI for further details), one obtains the pause-free elongation rate as

$$v = \frac{k_{\max}\,[\mathrm{NTP}]}{K\,[1+\sigma(F)]/\left(1+k_{\mathrm{TLc}}^+/k_{\mathrm{TLo}}^+\right)+[\mathrm{NTP}]}, \tag{4}$$

where $k_{\max} = k_{\mathrm{TLc}}^+/(1+k_{\mathrm{TLc}}^+/k_{\mathrm{TLo}}^+)$, $K = K_d + k_{\mathrm{TLc}}^+/k_b^0$, and in particular, $\sigma(F) \equiv \sigma_t(F) + \sigma_t(F)\sigma_{\mathrm{TLo}}(F)$, with $\sigma_{\mathrm{TLo}}(F) \equiv k_{\mathrm{TLo}}^-/k_{\mathrm{TLo}}^+$ and $\sigma_t(F) \equiv k_t^-/k_t^+$ as the biases *against* the TL opening (V → I) and translocation (I → II), respectively. The larger the bias, the system is more stabilized in the initial state (the TL closed state V or the pre-translocated state I). It is easy to see that equation (4) is analogous to equation (1), derived for the non-branched BR model in the three-state representation. Mathematically, fittings to experimental data according to equations (4) and (1) give equivalent results for the dominant or slow events: One obtains $k_{\mathrm{TLc}}^+ \sim 35\ \mathrm{s}^{-1}$ and $k_{\mathrm{TLo}}^+ \sim 112\ \mathrm{s}^{-1}$ (under the assisting force of 6.5 pN) in model A.

Combing with experimental data on pause densities [32], one can estimate quantities such as K, but cannot extract further the 'intrinsic' information K_d and k_b^0 on NTP binding. One also notices that in equation (4) the elongation velocity is independent of some of individual kinetic rates, such as k_{TLc}^- and k_c^+, as the derivation is under specific assumptions of model A. Fitting equation (4) to the elongation velocity data, therefore, cannot reveal those 'hidden' parameters. In practice, one can also fit the experimental data following the exact formulae (e.g., see SI equation S1) when the model assumptions hold approximately. As a result, some parameters can be estimated this way while the rest are assigned to likely values. The fitting details of model A to the experimental data are provided in SI figure S2 and table S1.

In the most recent MD simulation study that demonstrated the fast translocation of Pol II, it was found that the forward rate and backward rate are almost identical, at $\sim 5\times10^4\ \mathrm{s}^{-1}$ [48]. Accordingly, we set $\sigma_t(0)=1$ at the zero force. To keep the results consistent with the measured pause densities of Pol II during elongation [35], one obtains $K \sim \frac{9.2}{1+\sigma_{t\,(F=6.5\ \mathrm{pN})}} \sim 5.8$ μM, and $k_{\mathrm{TLo}}^- \sim 2142\ \mathrm{s}^{-1}$ at zero force, thus the bias against the TL opening is $\sigma_{\mathrm{TLo}}(0) \sim 21$ (see SI table S1). It indicates that prior to translocation, the TL-closed product state (V) is more favored than the TL open pre-translocation state (I), by $k_{\mathrm{B}}T\ln 21 \sim 3\,k_{\mathrm{B}}T$.

In estimating the pause densities, the pause related backtracking is assumed to start from state V, before the TL opening. Since the experimentally fitted backtracking rate is at $\sim 7\ \mathrm{s}^{-1}$, it presumably includes a slow process such as the TL opening. If the backtracking is assumed to start at the pre-translocation state I, one either cannot fit with the pause

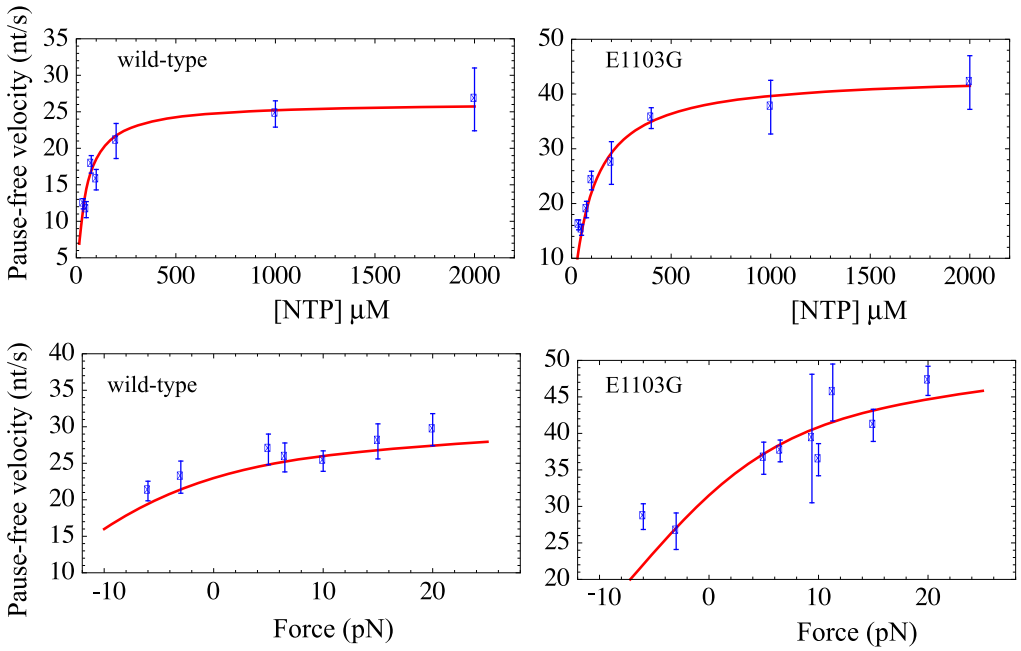


Figure 3. Fitting of model B to experimentally measured pause-free elongation velocities. The pause-free velocity versus the NTP concentration (upper row; at an assisting force $F = 6.5$ pN) and versus the force implemented (bottom; at [NTP] of 1 mM), for the wild-type Pol II (left) and the mutant polymerase E1103G (right) are shown, respectively. The blue data points were obtained from single molecule experiments [34, 35]. The red curve is the model fitting to the experimental data. The parameters and details are provided in SI (table S2). Similar fitting of model A can be found in SI figure S2 and table S1.

density data (at $\sigma_t(0) = 1$), or need put a heavy bias toward the post-translocated state II (see SI).

Essentially, one sees that the population bias toward the pre-translocation state (~7.7) presented in the three-state BR model [35] is now 'shifted' to the TL opening process prior to the translocation (~21), favoring the TL-closed product state (V). Furthermore, one infers that the intrinsic NTP dissociation constant $K_d < K = K_d + k^+_{\mathrm{TLc}}/k^0_b \sim 5.8$ μM, and the NTP binding rate constant estimated as $k^0_b > 6$ $\mu\mathrm{M}^{-1}$ s^{-1} (as $k^+_{\mathrm{TLc}}/k^0_b < 5.8$ μM). Note that one still has $K_M \equiv K[1 + \sigma(F)]/(1 + k^+_{\mathrm{TLc}}/k^+_{\mathrm{TLo}}) \sim 39$ μM, as measured experimentally [35].

From previous literature, the measured apparent NTP dissociation constant of RNA polymerases (or K_M in the Michaelis–Menten scheme) ranges from 10 to 100 μM [51–53]. Conventionally, one assumes that the intrinsic dissociation constant K_d is similar to the apparent value. Model A indicates that these values can be significantly different in the polymerase elongation kinetics. While K_M was measured ~39 μM, $K < 10$ μM is obtained, and K_d is even smaller. For example, in table S1, when we set k^0_b=20 $\mu\mathrm{M}^{-1}\,\mathrm{s}^{-1}$, K_d is ~4.1 μM. In an alternative model B, however, we show that K_d is not restricted to such a low value.

2.5. Model B: the five-state BR model with a rate-limiting step at catalytic stage after TL closing

In model B, we keep all above settings from model A but change only one feature: transition IV → V (forward rate k^+_c, without force dependence) becomes rate limiting, instead of III → IV rate limiting in model A. The pause-free elongation rate can be written as:

$$v = \frac{k_{\max}[\mathrm{NTP}]}{K[1 + \sigma(F)]/\left(1 + k^\star_c/k^+_{\mathrm{TLo}}\right) + [\mathrm{NTP}]}, \quad (5)$$

where $k_{\max} = k^\star_c/(1 + k^\star_c/k^+_{\mathrm{TLo}})$, $K = K_d\frac{\sigma_{\mathrm{TLc}}}{1+\sigma_{\mathrm{TLc}}} + k^\star_c/k^0_b$, in which $k^\star_c \equiv k^+_c/(1 + \sigma_{\mathrm{TLc}})$, and $\sigma_{\mathrm{TLc}} \equiv k^-_{\mathrm{TLc}}/k^+_{\mathrm{TLc}}$, the bias against TL closing upon NTP binding. Still, one keeps the force-dependent biasing factor as that in equation (4): $\sigma(F) \equiv \sigma_t(F) + \sigma_t(F)\sigma_{\mathrm{TLo}}(F)$. Since the catalytic rate $k^+_c = k^\star_c(1 + \sigma_{\mathrm{TLc}})$ is assumed low in model B, while $k^\star_c$ needs to fit with the slowest rate in the experiment [35], σ_{TLc} cannot take a large value.

According to data fitting, $k^\star_c \sim 35$ s^{-1} and $k^+_{\mathrm{TLo}} \sim 112$ s^{-1} (at $F = 6.5$ pN) apply for model B as the two slow rates. $K \sim \frac{9.2}{1+\sigma_{t\,(F=6.5\ \mathrm{pN})}} \sim 5.8$ μM and $k^-_{\mathrm{TLo}} \sim 2142$ s^{-1} (at zero force, or $\sigma_{\mathrm{TLo}}(0) \sim 21$, see SI table S2) are obtained similarly as that in model A. A significant difference of model B from A is that $K = K_d\frac{\sigma_{\mathrm{TLc}}}{1+\sigma_{\mathrm{TLc}}} + k^\star_c/k^0_b$, such that $K_d < 5.8\frac{1+\sigma_{\mathrm{TLc}}}{\sigma_{\mathrm{TLc}}}$ μM. In particular, if the TL closing step is close to irreversible, i.e., $\sigma_{\mathrm{TLc}} \ll 1$ (or $k^-_{\mathrm{TLc}} << k^+_{\mathrm{TLc}}$), the intrinsic NTP dissociation constant can become quite large. In SI table S2, we set $\sigma_{\mathrm{TLc}} \sim 0.1$ and $k^0_T \sim 20$ $\mu\mathrm{M}^{-1}$ s^{-1}, thus $K_d \sim 45$ μM, which is significantly larger than $K \sim 5.8$ μM and becomes comparable to the measured $K_M \sim 39$ μM [35]. One can see the measurements along with the model B data fitting in figure 3, where the pause-free elongation rate of wild-type Poll II

approaches ~25 nt s^{-1} (under the force of 6.5 pN) with increasing concentrations of NTP (~39 μM). Besides, the force dependence of the elongation rate (at a high NTP concentration of 1 mM) is also demonstrated: variation of the forces from opposing to assisting (from −6 pN to 20 pN) increases the elongation rate from ~20 nt s^{-1} to ~30 nt s^{-1}. The measurements and data fitting for the mutant E1103 are also shown in figure 3.

By analyzing and comparing model A and B, we see that a generic non-branched BR mechanism explains the experimental data sufficiently well. In both models, the *force-dependent* slow step can be attributed to the TL opening that precedes the fast translocation; NTP binding is only allowed at post-translocation, acting as a paw for the BR. Model A assigns the *force-independent* slow step to the TL closing or isomerization transition (III → IV) right upon NTP binding. In this case, the polymerase waits a long time at state III for transiting to state IV, thus state III is highly populated. As such, one expects a fairly small NTP dissociation constant so that NTP is held tightly at state III to avoid excessive dissociation. In contrast, model B assigns the slow step to the catalytic transition (IV → V, but dominated by pre-catalytic adjustments), while the TL closing isomerization becomes relatively fast (e.g. at a rate >1000 s^{-1}). In case that the TL closing upon NTP binding is close to irreversible, state III becomes transient or lowly populated. Accordingly, the NTP dissociation constant can be large, as the polymerase does not lose NTP significantly from the lowly populated state III.

One can also compare the fitting results of the mutant E1103G with that of the wild type, for both model A and B. According to the measurements [35], one can estimate quantitative features of the mutant polymerase at a fast rate limit (e.g. the force-independent rate-limiting step proceeds as fast as ~260 s^{-1}, see SI for details). From the experimental fitting, the forward rate of force-dependent step ($k^{+}_{\mathrm{TLo}}$for the TL opening prior to translocation) decreases ~50% in the mutant, in comparison to the wild type. The bias against the TL opening is found ~20 in the wild type (for both model A or B), while it increases to at least ~30 in the mutant (with $K \sim 60/(1+\sigma_t)$ μM set for the mutant, see SI). Hence, one is able to attribute the mutant impact on the force-dependent step to stabilization of the TL closed configuration in the product state, and this impact is similar in both model A and B. On the other hand, the rate of the force-independent step increases significantly in the mutant.

As the bias against the force-dependent TL opening is obtained above, one gets the free energy of the TL closed product state (V) ~ 3 k_BT (or ~2 kcal mol^{-1}) lower than that of the TL open pre-trans state (I). This free energy difference matches well to that has been estimated from a recent MD simulation study [26]. The study also estimated that the free energy barrier from the TL open to the closed state is about 2–4 k_BT (1.5–2.5 kcal mol^{-1}) lower than that from the TL closed to the open state, in the presence of NTP. The property seems consistent with model B, in which the TL closed to the open (V → I) happens at ~100 s^{-1}, while the TL open to the closed (III → IV) happens about tens of times faster (~1000 s^{-1} or over). In model A, the TL closing upon NTP binding is even more slowly than the TL opening prior to translocation. Anyhow, we note that the barriers reported in [26] were obtained using a combination of target MD and Hamiltonian replica exchange MD simulations, in which the application of the external force may introduce artifacts.

3. Discussion

Pol II is a complex molecular machine that cycles though a number of conformational states upon each nucleotide addition. Although various intermediate states have been characterized by biochemical and structural experiments, the dynamical mechanisms are still lack of. By combing existing biochemical, structural, single-molecule, and MD simulations studies, we develop a transcription elongation model for Pol II in which a slow transition of TL opening happens right before translocation of the polymerase. Overall, the elongation still follows a generic BR mechanism, with NTP binding after the fast Brownian translocation. Essentially, our model suggests that the TL opening instead of the translocation brings the force-dependent slow transition detected in the single molecule experiment [35]. The proposal is supported by structure-based computational studies: previous MD simulation and normal mode analyses based on molecular force field had indicated that the TL in its closed form inhibits the downstream DNA translocation, so that opening of the TL is necessary to remove the inhibition [41, 42]. Recent MD simulation connecting a large number of short trajectories to a network of Markov states demonstrated that the translocation of Pol II happens at ~20 μs [48], which is much faster than the milliseconds single molecule measurements. The simulated translocation, however, does not involve TL opening transition, and only contains a minimum scaffold of transcription complex [54]. In addition, MD simulation studies had also shown that the TL only slightly opens after a fast PPi release at microseconds [43, 44], while the substantial opening of the TL is left to happen thereafter.

The mutant Pol II E1103G demonstrates consistent behaviors with this TL opening model. We notice that E1103 locates at one end of the TL. It forms a suboptimal salt bridge with R1100 that likely destabilizes the local helix folding in the closed/folded TL [49]. The mutation E1103G abolishes the destabilization effect of the E1103-R1100 salt bridge and thus stabilizes the TL in its folded/closed configuration. Accordingly, the mutant E1103G exhibits a slower rate

of TL opening, a stronger bias against the TL opening, and in addition, a faster rate of TL closing or catalysis, comparing to the wild type Pol II [35].

On the other hand, previous work combining single experiments with kinetic modeling demonstrated sequence-dependent pausing behaviors of the polymerase during elongation [55]. Due to variation of the sequence stabilities along the double stranded DNA, each step of the polymerase translocation brings a variable free energy change $\Delta G = G_{\mathrm{post-trans}} - G_{\mathrm{pre-trans}}$, caused by melting of the downstream base pair (bp), re-annealing of the upstream bp, and similar adjustments in the DNA–RNA hybrid [55, 56]. Correspondingly, the translocation bias varies as $\sigma_t \equiv \frac{k_t^-}{k_t^+} = \sigma_t(0)\exp(\Delta G/k_{\mathrm{B}}T)$, so that $K_M = K[1+\sigma_t]/(1+k_c/k_t^+)$ in the pause-free velocity (e.g. equation (1)) incorporates the sequence effect. Indeed, a sequence barrier $\Delta G > 0$ above thermal fluctuation level can lead to a very large K_M, and thus interfere with the overall elongation rate. In SI, we numerically examined if it is possible that the force-dependent slowing down detected in the single molecule experiment [35] is indeed caused by the sequence barriers. Our calculation shows that only when large sequence barriers ($\Delta G \sim 5$ to 6 $k_{\mathrm{B}}T$ or over) are populated sufficiently high (e.g. 10–30% over), can the sequence barriers slow down the elongation rate significantly. For regular sequences with ΔG varying within $\sim 3\ k_{\mathrm{B}}T$, the sequence barriers cannot cause that much force-dependent slowing down as detected. As such, it becomes even convincing that it is the slow conformational transition (i.e., TL opening) susceptible to mechanical force change that causes the force-dependent slowing down in the single molecule detection.

To further explore the kinetic specificities in the Pol II elongation, we have analyzed two possible scenarios for the rate-limiting event *after* NTP binding. The first scenario, presented in model A, assumes that a slow event starts right after NTP binding or pre-insertion. This slow event delivers a substantial isomerization or TL closing transition during NTP insertion. Correspondingly, the NTP-bound 'pre-insertion' state (III) becomes highly populated as the initial state of the transition. During the long waiting period of the slow transition, the NTP affinity to the binding site has to be sufficiently high in order to keep NTP from dissociation, or say, the NTP dissociation constant has to be sufficiently small. For the second scenario, presented in model B, the slowest transition happens one step further into the catalytic stage. In this case, the NTP binding affinity can be either high or low, as the isomerization transition upon NTP binding proceeds sufficiently fast to allow a timely NTP insertion (e.g. $\sim 1000\ \mathrm{s}^{-1}$ or over). Since both the phosphoryl transfer reaction and PPi release happen fast, the slow transition is attributed to some pre-catalytic adjustment, such as rearrangement or tightening of the active site [21, 42]. The rearrangement assists a proper geometry of NTP to form in the active site, so that an efficient phosphoryl transfer reaction can happen. In order to determine which scenario is more likely, one can examine the *intrinsic* NTP dissociation constant $K_d \equiv k_b^-/k_b^0$: according to current analyses, if $K_d > \sim 10\ \mu$M, model A would be ruled out.

In common practice, however, it is the apparent value of the NTP dissociation constant or indeed, the Michaelis constant K_M that is measured, instead of the intrinsic value. Nevertheless, even in the simple three-state BR scheme, the derivation based on the master equation approach shows the discrepancy between $K_M = K(1+\sigma_t)/(1+k_c^+/k_t^+)$ (see equation (1); $K \equiv K_d + k_c^+/k_b^0$) and K_d: only when $\sigma_t \ll 1$ (irreversible translocation) and $k_c^+ \ll k_t^+$ (very slow catalysis), $K_M = K$; and if $k_c^+/k_b^0 \ll K_d$(very slow catalysis), then $K_M = K = K_d$. From literature search, however, we found that most measured NTP dissociation constants are indeed K_M (as assumed in the conventionally simplified Michaelis–Menten kinetics), and the values range from $\sim 10\ \mu$M to 100 μM or higher [51–53, 57]. As denoted above and from [55], K_M is highly sequence-dependent as $\sigma_t \propto \exp(\Delta G/k_{\mathrm{B}}T)$. Hence, it is reasonable to see that a range of values of K_M had been obtained experimentally. In particular, Bai *et al* [57] closely measured K (denoted K_d^{NTP} in table 1 of [57]; note that $K = K_d$ when the catalysis is much slower than the translocation) for different species of RNA nucleotides in *E. coli* RNAP elongation. For CTP, $K \sim 9\ \mu$M and $k_c^+ \sim 43\ \mathrm{s}^{-1}$ were obtained, so that one infers that $k_c^+/k_b^0 < 9\ \mu$M for CTP and $k_b^0 > \sim 5\ \mu\mathrm{M}^{-1}\ \mathrm{s}^{-1}$ for all nucleotides (k_b^0 assumed independent of NTP species). Similarly, one infers that $K_d > 36\ \mu$M, 25 μM, and 72 μM for ATP, UTP and GTP (see SI). If one assumes that the intrinsic NTP dissociation properties are similar for Pol II and the *E. coli* RNAP [58], then it is likely an average K_d will be substantially larger than 10 μM in Pol II. In that case, model A would be ruled out, and model B would be supported. There is some previous evidence also suggested that the TL closing transition after NTP binding happens fast [59], as supported by model B, in which the force-independent rate-limiting step is set at the catalytic stage. Hence, one sees that accurate measurements of the intrinsic NTP dissociation constants or NTP binding affinities help essentially on discovering a more specific elongation scenario.

Indeed, the specific details of NTP binding to Pol II remains to be elusive and controversial. It had been proposed that there is an entry (E-) site aside of the active (A-) site for the initial NTP binding [52, 60–62]. One previous modeling work on diffusion of NTP into the polymerase active center showed that binding to the E-site adjacent to the A-site can overcome the limitation of the RNA synthesis at low NTP concentration [61]. The NTP

binding had also been suggested to happen early in the pre-translocation state [60]. The branched BR model adopts this perspective, allowing NTP binding at either pre- or post-translocation [34, 37]. The branched BR model provides a fairly large upper bound to the intrinsic NTP dissociation constant ($K_d < 140\ \mu$M), as an extra site greatly enhances the life time of NTP around the active center [61]. Superimposing of NTP into the E-site of the pre-translocated structure of Pol II (see figure 2(b)) shows a high proximity between the NTP triphosphate and the backbone phosphate of the 3'-end of the RNA. Normally, due to the strong electrostatic repulsion, the two negatively charged groups would not be able to approach each other to such a short distance (~3.0 Å between two closest oxygen atoms). Nevertheless, our docking study indicates that it is possible for the substrate NTP to directly bind to the E-site of the pre-translocation state of Pol II (figure 2(b)). In particular, several positively charged residues (Lys987, Arg766 and Arg1020), along with a magnesium ion in the active site, help on shielding the negative electrostatic repulsions. Therefore, the branched BR model that allows NTP binding at pre-translocation is still feasible. Nevertheless, the fitting results of the branched BR model do not seem to match well with current computational findings. For example, the translocation bias $\sigma_t^0 \sim 0.2$ suggested in the branched BR model has not been identified in the MD studies [48]. It is also not clear whether TL opening can proceed fast prior to the translocation, and how TL opening and NTP binding are possibly coordinated at the pre-translocation. Hence, we adopt only the non-branched BR mechanism in current study.

In addition, next to the TL structure around the active center, one can identify an essential BH. From the simulation study, the BH bends frequently during translocation of Pol II [48]. Close examination shows that there is a high chance that the BH bending interferes with the active center and hence the A-site NTP binding: the middle region of the BH would bend into the A-site for ~70% chance at pre-translocation, and for ~50% chance during the translocation (see SI for details). Only at post-translocation, the BH does not bend that much, so it hardly interferes with the active center or NTP binding to the A-site any more. Accordingly, one should be aware that even though NTP binding to the E-site is plausible at pre-translocation, the binding to the A-site is not fully supported until the post-translocation state is reached. Interestingly, one can also compare the multi-subunit Pol II translocation with that of the single-subunit T7 RNA polymerase [63], in which a highly conserved residue Tyr639 oscillates IN and OUT the active site similarly as the BH bending region, while an O-helix opens and closes through the elongation cycle as TL does in Pol II.

4. Conclusions

Based mainly on most recent single molecule experimental measurements and computational studies on Pol II transcription elongation, we present a kinetic model that connects the experimental data analyses with the structure-based modeling and simulation. The model follows a generic BR mechanism, in which the polymerase translocates forward and backward spontaneously and quickly, until NTP binds at post-translocation to stop the backward translocation. Since the single molecule measurements suggested a slow force-dependent transition prior to NTP binding, we show that the slow transition can be a TL opening process that precedes the translocation, while the translocation per se is still fast. Consistently, one can see that the mutant polymerase E1103G stabilizes the closed TL configuration and also slows down the TL opening prior to the translocation. In addition, we consider the sequence-dependent translocation and show that the significant force-dependent slowing down of the Pol II elongation, as detected experimentally, cannot be attributed to the regular sequence stability variation along DNA. On the other hand, in current model, it is still to be determined which step after NTP binding is rate limiting for regular cognate NTP incorporation. According to our analyses, if this rate-limiting step happens right after NTP binding, as during an isomerization for the TL closing, then the NTP binding affinity has to be high to prevent excessive NTP dissociation. Consequently, a low NTP binding affinity or high NTP dissociation constant indicates that instead, the force-independent rate-limiting event happens later, after TL closing, as for the pre-catalytic arrangements. Though current study follows a generic non-branched BR mechanism, one cannot rule out the branched BR model, as it seems plausible that NTP binds to the E-site prior to the translocation. As such, we emphasize that accurate determination of NTP binding properties, either quantitatively on the NTP affinity or dissociation constant, or qualitatively on when and where NTP binds to the polymerase elongation complex, would substantially improve our knowledge and understanding on the polymerase elongation. It is notable that the working model developed here not only provides structural and functional insights in Pol II transcriptional elongation mechanisms, but also makes critical connections between the overall elongation kinetics with individual transition steps, which can be characterized relatively easy by existing experimental and computational techniques. Hence, the modeling perspective brings new opportunities of investigating how local structural and dynamical perturbations may affect overall elongation kinetics, for example, on how a DNA damage that hinders translocation may ultimately affect the *in vitro* elongation rate.

Acknowledgments

We are grateful to Bustamante lab from Berkeley and Block lab from Stanford for sharing their single molecule experimental data for this study. We thank Dr Toyotaka Ishibashi, Dr Michelle Wang, and Dr Dong Wang for helpful discussions with us. We acknowledge the support from National Science Foundation of China Grant (no. 11275022 to JY and no. 21273188 to XH), and Hong Kong Research Grants Council GRF 16302214, ECS 60981, and AoE/M-09/12 to XH.

References

[1] Young R A 1991 RNA polymerase II *Ann. Rev. Biochem.* **60** 689–715
[2] Kostek S A, Grob P, De Carlo S, Lipscomb J S, Garczarek F and Nogales E 2006 Molecular architecture and conformational flexibility of human RNA polymerase II *Structure* **14** 1691–700
[3] Kornberg R D 2007 The molecular basis of eukaryotic transcription *Proc. Natl Acad. Sci. USA* **104** 12955–61
[4] Svetlov V and Nudler E 2013 Basic mechanism of transcription by RNA polymerase II *Biochimica Biophysica Acta* **1829** 20–8
[5] Kornberg R D 1999 Eukaryotic transcriptional control *Trends Cell Biol.* **9** M46–9
[6] Sims R J III, Belotserkovskaya R and Reinberg D 2004 Elongation by RNA polymerase II: the short and long of it *Genes Dev.* **18** 2437–68
[7] Greive S J and von Hippel P H 2005 Thinking quantitatively about transcriptional regulation *Nat. Rev. Mol. Cell Biol.* **6** 221–32
[8] Fuda N J, Ardehali M B and Lis J T 2009 Defining mechanisms that regulate RNA polymerase II transcription *in vivo Nature* **461** 186–92
[9] Zhou Q, Li T and Price D H 2012 RNA polymerase II elongation control *Ann. Rev. Biochem.* **81** 119–43
[10] He Y, Fang J, Taatjes D J and Nogales E 2013 Structural visualization of key steps in human transcription initiation *Nature* **495** 481–6
[11] Cramer P, Bushnell D A and Kornberg R D 2001 Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution *Science* **292** 1863–76
[12] Gnatt A L, Cramer P, Fu J, Bushnell D A and Kornberg R D 2001 Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution *Science* **292** 1876–82
[13] Woychlk N A and Hampsey M 2002 The RNA polymerase II machinery: structue illuminates function *Cell* **108** 453–63
[14] Hahn S 2004 Structure and mechanism of the RNA polymerase II transcription machinery *Nat. Struct. Mol. Biol.* **11** 394–403
[15] Wang D, Bushnell D A, Westover K D, Kaplan C D and Kornberg R 2006 Structural basis for transcription: role of the trigger loop in substrate specificty and catalysis *Cell* **127** 941–54
[16] Brueckner F, Ortiz J and Cramer P 2009 A movie of the RNA polymerase nucleotide addition cycle *Curr. Opin. Struct. Biol.* **19** 294–9
[17] Wang D, Bushnell D A, Huang X, Westover K D, Levitt M and Kornberg R D 2009 Structural basis of transcription: back-tracked RNA polymerase II at 3.4 angstrom resolution *Science* **324** 1203–6
[18] Kostrewa D, Zeller M E, Armache K-J, Seizl M, Leike K, Thomm M and Cramer P 2009 RNA polymerase II–TFIIB structure and mechanism of transcription initiation *Nature* **462** 323–30
[19] Liu X, Bushnell D A, Wang D, Calero G and Kornberg R D 2010 Structure of an RNA polymerase II–TFIIB complex and the transcription initiation mechanism *Science* **327** 206–9
[20] Martinez-Rucobo F W and Cramer P 2013 Structural basis of transcription elongation *Biochimica Biophysica Acta* **1829** 9–19
[21] Toulokhonov I, Zhang J, Palangat M and Landick R 2007 A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing *Mol. Cell* **27** 406–19
[22] Nudler E 2009 RNA polymerase active center: the molecular engine of transcription *Ann. Rev. Biochem.* **78** 335–61
[23] Huang X, Wang D, Weiss D R, Bushnell D A, Kornberg R D and Levitt M 2010 RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription *Proc. Natl Acad. Sci. USA* **107** 15745–50
[24] Zhang J, Palagnat M and Landick R 2010 Role of the RNA polymerase trigger loop in catalysis and pausing *Nat. Struct. Mol. Biol.* **17** 99–104
[25] Fouqueau T, Zeller M E, Cheung A C, Cramer P and Thomm M 2013 The RNA polymerase trigger loop functions in all three phases of the transcription cycle *Nucleic Acids Res.* **41** 7048–59
[26] Wang B, Predeus A V and Burton Z F 2013 Energetic and structural details of the trigger-loop closing transition in RNA polymerase II *Biophys. J.* **105** 767–75
[27] Hein P P and Landick R 2010 The bridge helix coordinates movements of modules in RNA polymerase *BMC Biol.* **8** 141–4
[28] Weinzierl R O J 2011 The bridge helix of RNA polymerase acts as a central nanomechanical switchboard for coordinating catalysis and substrate movement *Archaea* **2011** 608385
[29] Herbert K M, Greenleaf W J and Block S M 2008 Single-molecule studies of RNA polymerase: motoring along *Ann. Rev. Biochem.* **77** 149–76
[30] Selth L A, Sugurdsson S and Svejstrup J Q 2010 Transcript elongation by RNA polymerase II *Ann. Rev. Biochem.* **79** 271–93
[31] Kireeva M, Kashlev M and Burton Z F 2010 Translocation by multi-subunit RNA polymerases *Biochimica Biophysica Acta* **1799** 389–401
[32] Kaplan C D 2013 Basic mechanisms of RNA polymerase II activity and alteration of gene expression in *Saccharomyces cerevisiae Biochimica Biophysica Acta* **1829** 39–54
[33] Zhang S and Wang D 2013 Understanding the molecular basis of RNA polymerase II transcription *Isr. J. Chem.* **53** 442–9
[34] Larson M H, Zhou J, Kaplan C D, Palangat M, Kornberg R D, Landick R and Block S M 2012 Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II *Proc. Natl Acad. Sci. USA* **109** 6555–60
[35] Dangkulwanich M, Ishibashi T, Liu S, Kireeva M L, Lubkowska L, Kashlev M and Bustamante C 2013 Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism *eLIFE* **2** e00971
[36] Dangkulwanich M, Ishibashi T, Bintu L and Bustamante C 2014 Molecular mechanisms of transcription through single-molecule experiments *Chem. Rev.* **114** 3203–23
[37] Abbondanzieri E A, Greenleaf W J, Shaevitz J W, Landick R and Block S M 2005 Direct observation of base-pair stepping by RNA polymerase *Nature* **438** 460–5
[38] Bar-Nahum G, Epshtein V, Ruckenstein A E, Rafikov R, Mustaev A and Nudler E 2005 A ratchet mechanism of transcription elongation and its control *Cell* **120** 183–93
[39] Peskin C S, Odell G M and Oster G F 1993 Cellular motions and thermal fluctuations: the Brownian ratchet *Biophys. J.* **65** 316–24
[40] Feig M and Burton Z F 2009 RNA polymerase II flexibility during translocation from normal mode analysis *Proteins* **78** 434–46
[41] Feig M and Burton Z F 2010 RNA polymerase II flexibility during translocation from normal mode analysis *Proteins* **78** 434–46
[42] Feig B and Burton Z F 2010 RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation *Biophys. J.* **99** 2577–86
[43] Da L, Wang D and Huang X 2011 Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II *J. Am. Chem. Soc.* **134** 2399–406

[44] Da L-T, Pardo Avila F, Wang D and Huang X 2013 A two-state model for the dynamics of the pyrophosphate ion release in bacterial RNA polymerase *PLoS Comput. Biol.* **9** e1003020
[45] Pardo-Avila F, Da L-T, Wang Y and Huang X 2013 Theoretical investigations on elucidating fundamental mechanisms of catalysis and dynamics involved in transcription by RNA polymerase *J. Theor. Comput. Chem.* **12** 1341005
[46] Kireeva M L, Opron K, Seibold S A, Domecq C, Cukier R I, Coulombe B, Kashlev M and Burton Z F 2012 Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase *BMC Biophys.* **5** 11–28
[47] Wang B, Feig M, Cukier R I and Burton Z F 2013 Computational simulation strategies for analysis of multisubunit RNA polymerases *Chem. Rev.* **113** 8546–66
[48] Silva D A, Weiss D R, Pardo Avila F, Da L T, Levitt M, Wang D and Huang X 2014 Millisecond dynamics of RNA polymerase II translocation at atomic resolution *Proc. Natl Acad. Sci. USA* **111** 7665–70
[49] Meuzelaar H, Tros M, Huerta-Viga A, van Dijk C N, Vreede J and Woutersen S 2014 Solvent-exposed salt bridges influence the kinetics of α-helix folding and unfolding *J. Phys. Chem. Lett.* **5** 900–4
[50] Vassylyev D G, Vassylyeva M N, Zhang J, Palangat M, Artsimovitch I and Landick I 2007 Structural basis for substrate loading in bacterial RNA polymerase *Nature* **448** 163–8
[51] Anand V S and Patel S S 2006 Transient state kinetics of transcription elongation by T7 RNA polymerase *J. Biol. Chem.* **281** 35677–85
[52] Foster J E, Holmes S F and Erie D A 2001 Allosteric binding of nucleotide triphosphate to RNA polymerase regulates transcription elongation *Cell* **106** 243–52
[53] Kellinger M W, Ulrich S, Chong J, Kool E T and Wang D 2012 Dissecting chemical interactions governing RNA polymerase II transcriptional fidelity *J. Am. Chem. Soc.* **134** 8231–40
[54] Imashimizu M and Kashelv M 2014 Unveiling translocation intermediates of RNA polymerase *Proc. Natl Acad. Sci. USA* **111** 7507–8
[55] Bai L, Shundrovsky A and Wang M D 2004 Sequence-dependent kinetic model for transcription elongation by RNA polymerase *J. Mol. Biol.* **344** 335–49
[56] Yager T D and Von Hippel P H 1991 A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli Biochemistry* **30** 1097–118
[57] Bai L, Fulbright R M and Wang M D 2007 Mechanochemical kinetics of transcription elongation *Phys. Rev. Lett.* **98** 068103
[58] Murakami K S 2013 X-ray crystal structure of *Escherichia coli* RNA polymerase σ70 holoenzyme *J. Biol. Chem.* **288** 9126–34
[59] Kireeva M L, Nedialkov Y A, Cremona G H, Purtov Y A, Lubkowska L, Malagon F, Burton Z F, Strathern J N and Kashlev M 2008 Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol. Cell* **30** 557–66
[60] Nedialkov Y A, Gong X Q, Hovde S L, Yamaguchi Y, Handa H, Geiger J H, Yan H and Burton Z F 2003 NTP-driven translocation by human RNA polymerase II. *J. Biol. Chem.* **278** 18303–12
[61] Batada N N, Westover K D, Bushnell D A, Levitt M and Kornberg R D 2004 Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center *Proc. Natl Acad. Sci. USA* **105** 17361–4
[62] Westover K D, Bushnell D A and Kornberg R D 2004 Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center *Cell* **119** 481–9
[63] Yu J and Oster G 2012 A small post-translocation energy bias aids nucleotide selection in T7 RNA polymerase transcription *Biophys. J.* **102** 532–41
[64] Morris G M, Huey R, Lindstrom W, Sanner M F, Belew R K, Goodsell D S and Olson A J 2009 Autodock4 and autodocktools4: automated docking with selective receptor flexibility *J. Comput. Chem.* **30** 2785–91

Supplementary Information to "Constructing Kinetic Models to Elucidate Structural Dynamics of a Complete RNA Polymerase II Elongation Cycle"

I. Fitting to single molecule experimental data according to the branched BR model and non-branched BR model A

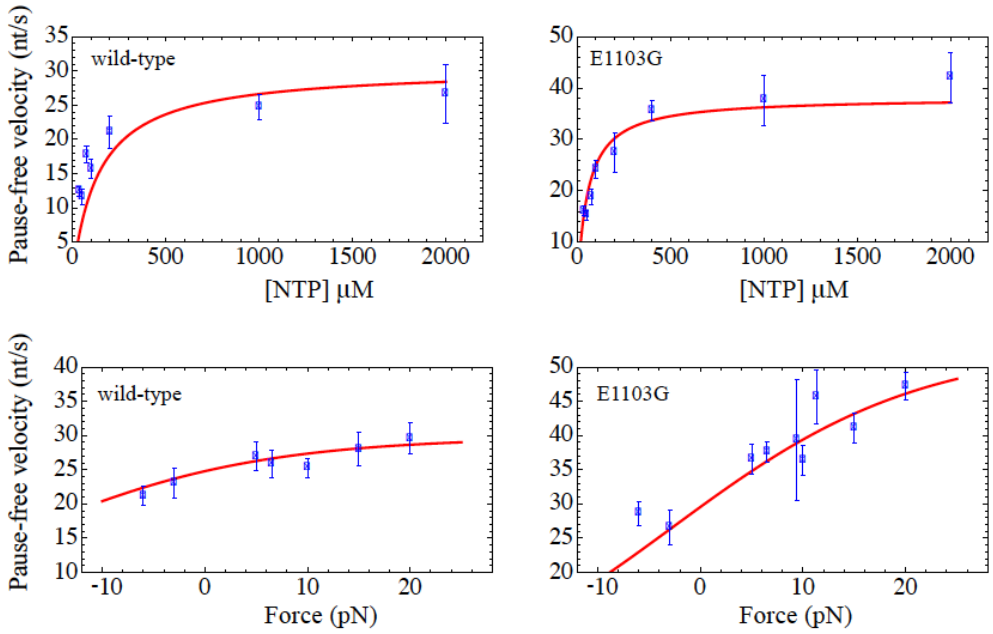


Fig S1 Plots of fitting to experimental data according to the branched Brownian ratchet mechanism (1). The fitting parameters are directly taken from the original work (Table 1 of the ref (1)). The experimental data (blue points) are collected from both ref (1) and (2). The original fitting was performed according to part of the force-velocity data from (1).

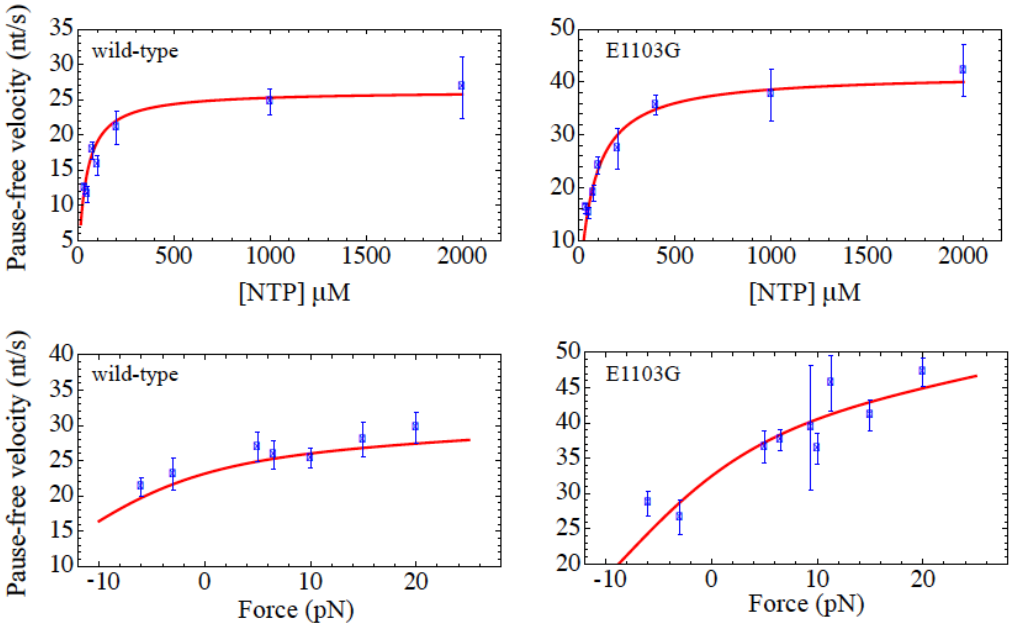


Fig S2 Fitting to experimental data according to **model A** in current work, following the non-branched generic Brownian ratchet mechanism. The experimental data (blue points) are collected from both ref (1) and (2). The fitting parameters are provided in **Table S1** below.

II. Derivation of model A and B following a generic non-branched BR mechanism

For the five-state elongation scheme depicted in **Fig 1c**, one can write down a master equation as that in main Eq (3). Usually, when the PPi concentration is low, the catalytic step is close to irreversible, $k_c^- \to 0$ is assumed so that to simplify the calculation. Nevertheless, when one calculates the chemical free energ $\Delta\mu$ of a cycle, one need count $\Delta\mu = k_B T(\ln\frac{k_t^+}{k_t^-} + \ln\frac{k_T^+}{k_T^-} + \ln\frac{k_{TLc}^+}{k_{TLc}^-} + \ln\frac{k_c^+}{k_c^-} + \ln\frac{k_{TLo}^+}{k_{TLo}^-})$ without using $k_c^- \to 0$. Note that at a steady state, $\frac{d}{dt}\Pi = 0$ is set for the master equation $\frac{d}{dt}\Pi = M\Pi$ (with $\Pi = (P_I\ P_{II}\ P_{III}\ P_{IV}\ P_{II})^T$). The probability flux J or the cycling rate is obtained as $J \equiv k_t^+ P_I - k_t^- P_{II} = ... = k_{TLo}^+ P_V - k_{TLo}^- P_I$.

II.1 Model A (a five-state BR model with TL closing rate-limiting)

$$J = \frac{k_{TLc}^+[NTP]}{[K_d(1+\frac{k_{TLc}^-}{k_c^+}) + k_{TLc}^+/k_b^0](1+\sigma_t+\sigma_t\sigma_{TLo}) + [1+\frac{k_{TLc}^+ + k_{TLc}^-}{k_c^+} + k_{TLc}^+/k_{TLo}^+(1+\frac{k_{TLo}^+ + k_{TLo}^-}{k_t^+})][NTP]}$$

$$\sim \frac{k_{TLc}^+/(1+k_{TLc}^+/k_{TLo}^+)[NTP]}{(K_d + k_{TLc}^+/k_b^0)(1+\sigma_t+\sigma_t\sigma_{TLo})/(1+k_{TLc}^+/k_{TLo}^+) + [NTP]}$$

(S1)

The above approximation is taken when $\frac{k_{TLc}^-}{k_c^+} \to 0$, $\frac{k_{TLc}^+ + k_{TLc}^-}{k_c^+} \to 0$, and $\frac{k_{TLo}^+ + k_{TLo}^-}{k_t^+} \to 0$. That is to say, the TL closing (and its reversal) after NTP binding happens much slower than the catalytic transition; and the TL opening (and its reversal) prior to the translocation happens much slower than the translocation. Under these approximations, one obtains Eq (4) in main.

Accordingly, one obtains $K_M \equiv K(1+\sigma)/(1+k_{TLc}^{+}/k_{TLo}^{+}) \sim 39$ μM as measured experimentally (2) (note $K \equiv K_d + k_{TLc}^{+}/k_b^0$ and $\sigma \equiv \sigma_t + \sigma_t\sigma_{TLo}$). Assuming that backtracking starts at product state V prior to the TL opening transition, then the experimental measurements on the pausing densities require that $\sigma_t k_{TLo}^{-} K \sim 4700$ μM s^{-1} (following similar analyses as in ref (2)). This way, one obtains $K \sim 9.2/(1+\sigma_t)$ and $k_{TLo}^{-} \sim \frac{4700(1+\sigma_t)}{9.2\sigma_t}$ in **model A**.

In the mutant E1103G, K_M = 62 ± 15 μM, $\sigma_t k_{TLo}^{-} K$ =(2.5 ± 0.4) x10^4 μM s^{-1}, k_{TLo}^{+}=50 ± 4 s^{-1}, and k_{TLc}^{+}=195 ± 65 s^{-1}. If one puts together all these average values, one cannot get a proper solution of $K > 0$. Hence, we only estimate the upper-bound value $K = K_{\max}$ by choosing $K_M \sim 77$ μM, $\sigma_t k_{TLo}^{-} K \sim 2.1$ x10^4 μM s^{-1}, $k_{TLo}^{+} \sim 54$ s^{-1} and $k_{TLc}^{+} \sim 260$ s^{-1} (as $K = \frac{K_M(k_{TLo}^{+}+k_{TLc}^{+}) - \sigma_t k_{TLo}^{-} K}{k_{TLo}^{+}(1+\sigma_t)}$). In that case, $K \sim 60/(1+\sigma_t)$ and $k_{TLo}^{-} \sim \frac{21000(1+\sigma_t)}{60\sigma_t}$ are obtained for the mutant. Note that these quantities are all considered under an assisting force F = 6.5 pN.

It is noted that if one assumes the backtracking starting at the pre-translocation I (after TL opening), then one obtains $k_t^{-} K \sim 4700$ μM s^{-1} so that to be consistent with the pause density measurements (2). With $k_t^{+} \sim$ 5x10^4 s^{-1}, $\sigma_t << 1$ has to hold (with $K \sim 51$ μM), which is at odds with the Brownian translocation.

II.2 Model B (a five-state BR model with catalytic step rate-limiting)

$$J = \frac{k_c^+[NTP]}{[K_d(\sigma_{TLc} + \frac{k_c^+}{k_{TLc}^+}) + k_c^+ / k_b^0](1+\sigma_t+\sigma_t\sigma_{TLo}) + [1+\sigma_{TLc} + \frac{k_c^+}{k_{TLc}^+} + k_c^+(1+\frac{k_{TLo}^+ + k_{TLo}^-}{k_t^+}) / k_{TLo}^+][NTP]}$$

$$= \frac{k_c^*[NTP]}{[K_d(\frac{\sigma_{TLc}}{1+\sigma_{TLc}} + \frac{k_c^*}{k_{TLc}^+}) + k_c^* / k_b^0](1+\sigma_t+\sigma_t\sigma_{TLo}) + [1+\frac{k_c^*}{k_{TLc}^+} + k_c^*(1+\frac{k_{TLo}^+ + k_{TLo}^-}{k_t^+}) / k_{TLo}^+][NTP]}$$

$$\sim \frac{k_c^* / (1+k_c^* / k_{TLo}^+)[NTP]}{(K_d \frac{\sigma_{TLc}}{1+\sigma_{TLc}} + k_c^* / k_b^0)(1+\sigma_t+\sigma_t\sigma_{TLo}) / (1+k_c^* / k_{TLo}^+) + [NTP]}$$

(S2)

The above approximation is taken as $\frac{k_c^*}{k_{TLc}^+} \to 0$ and $\frac{k_{TLo}^+ + k_{TLo}^-}{k_t^+} \to 0$, where $k_c^* \equiv \frac{k_c^+}{1+\sigma_{TLc}}$. That is to say, the catalysis (the pre-catalytic adjustment waiting) happens much slower than the TL closing transition right after NTP binding; and the TL opening (and its reversal) prior to the translocation happens much slower than the translocation. Under these approximations, one obtains Eq (5) in main. Note that the maximum rate becomes $k_{\max} = k_c^*$ in this case, and $K \equiv K_d\sigma_{TLc} / (1+\sigma_{TLc}) + k_c^* / k_b^0$. Except for these two quantities, other quantities and rest of the derivations are the same in **model A** and **B**.

From **Eq S1** and **S2**, one can see that the fittings made through **model A** and **B** are very similar to each other, if (*i*) the rate-limiting transition happens equally fast in both models: k_{TLc}^+ in **model A** equals k_c^* in **model B**, and (*ii*) the values of K are made identical in both models: K_d from **model A** equals $K_d\sigma_{TLc} / (1+\sigma_{TLc})$ from **model B**.

Parameter	Definition	Wild-type	E1103G	Notes
$k_{TLo}^{+6.5}$	Forward rate of the TL opening prior to translocation (I→II) at force F=6.5 pN	**112 s^{-1}** (k_{TLo}^{+} ~ 101 s^{-1} at zero force)	***< 54 s^{-1}*** (k_{TLo}^{+} ~ 51 s^{-1} at zero force)	Force-dependent $k_{TLo}^{+} = k_{TLo}^{+6.5} e^{(F-6.5)\delta_1 \Delta / k_B T}$ Δ = 1 nt distance see δ_1 below (it is set differently from that in (2)
$k_{TLo}^{-6.5}$	Backward rate of step of the above transition (II→I) at force F=6.5 pN	$\frac{4700}{9.2} \cdot \frac{1+\sigma_{t(F=6.5)}}{\sigma_{t(F=6.5)}}$ ~1392 s^{-1} (k_{TLo}^{-} ~ 2142 s^{-1} at zero force so that $\sigma_{TLo} \sim 21$)	$> \frac{21000}{60} \cdot \frac{1+\sigma_{t(F=6.5)}}{\sigma_{t(F=6.5)}}$ ~953 s^{-1} (k_{TLo}^{-} ~ 1548 s^{-1} at zero force so that $\sigma_{TLo} > 30$)	$k_{TLo}^{-} = k_{TLo}^{-6.5} e^{-(F-6.5)(1-\delta_1)\Delta / k_B T}$ See σ_t below
δ_1	Intermediate distance factor in the TL opening step $0 < \delta_1 < 1$ in $k_{TLo}^{\pm}$	*0.2± 0.2*	*0.1± 0.07*	Nonlinear fitting in Mathematica (c.f. Eq S1) *t-statistics: 1.2 and 1.6* *P-value: 0.3 and 0.15*
$k_{t(F=0)}^{+}$	Forward rate of the translocation at zero force	50000 s^{-1}	50000 s^{-1}	Estimated from MD simulation (3) $k_t^{+} = k_{t(F=0)}^{+} e^{F\delta_2 \Delta / k_B T}$ $k_t^{-} = k_{t(F=0)}^{-} e^{-F(1-\delta_2)\Delta / k_B T}$
σ_t^0	$\sigma_t^0 \equiv k_{t(F=0)}^{-} / k_{t(F=0}^{+}$ *translocation bias*	1.0	1.0	Set at 1.0 as for generic Brownian translocation $\sigma_t = \sigma_t^0 e^{-F\Delta / k_B T}$ ~ 0.58 at F=6.5 pN
δ_2	Intermediate distance factor in the translocation step: $0 < \delta_2 < 1$ in $k_t^{\pm}$	0.2	0.001	An arbitrary value does not affect velocity (c.f. Eq S1)
k_b^0	NTP binding constant such that $k_b^{+} = k_b^0 [NTP]$ for the forward rate of II→III	20 $\mu M^{-1} s^{-1}$	20 $\mu M^{-1} s^{-1}$	Set at this value $K_d = 9.2/(1+\sigma_{t(F=6.5)})$ $-35 / k_b^0$ ~4.1 μM for the wild type

				$K_d = 60/(1+\sigma_{t(F=6.5)})$ $-260 / k_b^0$ ~25 μM for the mutant
k_{TLc}^+	The rate of TL closing right after NTP biding	**35 s^{-1}**	~ **260 s^{-1}** (set at a large value)	The wild-type value is measured, while the mutant value is estimated and a rather large value is used
σ_{TLc}	$\sigma_{TLc} \equiv k_{TLc}^- / k_{TLc}^+$ *TL closing bias*	*0.1± 3.1*	*3.0 ± 1.8*	Fitting as k_c^+ is given below (c.f. Eq S1 exact form) *t-statistics: 0.03 1.7* *P-value: 0.98 0.15*
k_c^+	Forward rate of catalysis, which is assumed relatively large in model A	3000 s^{-1}	3000 s^{-1}	Set at the value
k_c^-	The reversal rate of the catalysis	1 s^{-1}	0.004	$\Delta\mu$ ~ 13 k_BT in both cases

Table S1 Parameters used in **model A**. The bold values of the parameters are obtained directly from the experimental study (2). The parameters that were set at certain values here are shaded (in gray). The values of $\sigma_{TLo}^{\pm}$ and K_d are highlighted (in blue and green).

Parameter	Definition	Wild-type	E1103G	Notes
$k_{TLo}^{+6.5}$	Forward rate of the TL opening prior to translocation (I→II) at force F=6.5 pN	**112 s^{-1}** (k_{TLo}^{+} ~ 101 s^{-1} at zero force)	***< 54 s^{-1}*** (k_{TLo}^{+} ~ 39 s^{-1} at zero force)	Force-dependent $k_{TLo}^{+} = k_{TLo}^{+6.5} e^{(F-6.5)\delta_1 \Delta / k_B T}$ Δ = 1 nt distance see δ_1 below
$k_{TLo}^{-6.5}$	Backward rate of the above transition (II→I) at force F=6.5 pN	$\frac{4700}{9.2} \cdot \frac{1+\sigma_{t(F=6.5)}}{\sigma_{t(F=6.5)}}$ ~1392 s^{-1} (k_{TLo}^{-} ~ 2142 s^{-1} at zero force so that $\sigma_{TLo} \sim 21$)	$> \frac{21000}{60} \cdot \frac{1+\sigma_{t(F=6.5)}}{\sigma_{t(F=6.5)}}$ ~953 s^{-1} (k_{TLo}^{-} > 1582 s^{-1} at zero force so that $\sigma_{TLo} > 40$)	$k_{TLo}^{-} = k_{TLo}^{-6.5} e^{-(F-6.5)(1-\delta_1)\Delta / k_B T}$ see σ_t below and illustration later
δ_1	Intermediate distance factor in the TL opening step: $0 < \delta_1 < 1$ in $k_{TLo}^{\pm}$	*0.2 ± 0.2*	*0.06 ± 0.06*	Nonlinear fitting in Mathematica (c.f. Eq S2) *t-statistics: 1.2 and 0.9* *P-value: 0.3 and 0.4*
$k_{t(F=0)}^{+}$	Forward rate of the translocation at zero force	50000 s^{-1}	50000 s^{-1}	Estimated from MD simulation(3) $k_t^{+} = k_{t(F=0)}^{+} e^{F\delta_2 \Delta / k_B T}$ $k_t^{-} = k_{t(F=0)}^{-} e^{-F(1-\delta_2)\Delta / k_B T}$
σ_t^0	$\sigma_t^0 \equiv k_{t(F=0)}^{-} / k_{t(F=0)}^{+}$ *translocation bias*	1.0	1.0	Set at 1.0 as for a standard Brownian translocation $\sigma_t = \sigma_t^0 e^{-F\Delta / k_B T}$ ~ 0.58 at F=6.5 pN
δ_2	Intermediate distance factor in the translocation: $0 < \delta_2 < 1$ in $k_t^{\pm}$	0.2	0.001	An arbitrary value does not affect velocity (c.f. Eq S2)
k_b^0	NTP binding constant such that $k_b^{+} = k_b^0 [NTP]$ for the forward rate of II->III	20 $\mu M^{-1} s^{-1}$	20 $\mu M^{-1} s^{-1}$	Set at this value $K_d = [9.2/(1+\sigma_{t(F=6.5)}) - 35 / k_b^0] \frac{1+\sigma_{TLc}}{\sigma_{TLc}}$ ~45 µM for the wild type; $K_d = [60/(1+\sigma_{t(F=6.5)}) - 260 / k_b^0] \frac{1+\sigma_{TLc}}{\sigma_{TLc}}$ ~ 302 µM for the mutant

k_{TLc}^{+}	The rate of TL closing right after NTP biding	3000 s^{-1}	3000 s^{-1}	Set at the value
σ_{TLc}	$\sigma_{TLc} \equiv k_{TLc}^{-} / k_{TLc}^{+}$ *TL closing bias*	0.1	*0.09± 0.05*	Set the value for the wild type (cannot fit well, c.f. Eq S2 approximate form; as $\frac{k_c^*}{k_{TLc}^+} \to 0$, σ_{TLc} keeps 'hidden'). Fitting as k_{TLc}^{+} is given above for the mutant (c.f. Eq S2 exact form as the approximation is not good) *t-statistics: 1.76* *P-value: 0.12*
k_c^+	The catalytic rate, which is supposed to be small in model B	**35 (1+ σ_{TLc}) ~s^{-1}**	***< 260* (1+ σ_{TLc}) *s^{-1}***	*a large value estimated for the mutant*
k_c^-	The reversal rate of the catalysis	0.0016 s^{-1}	0.002 s^{-1}	$\Delta\mu$ ~ 13 k_BT in both cases

Table S2 Parameters used in **model B**. The bold values of the parameters are obtained directly from the experimental study (2). The parameters that were set at certain values here are shaded (in gray). The values of $\sigma_{TLo}^{\pm}$ and K_d are highlighted (in blue and green).

III Sequence effect on translocation slowing down

Eq (1) in main describes the pause-free velocities of the polymerase elongation, in the three-state kinetic model under a generic BR mechanism. The forward and backward translocation rates depend on an assisting or opposing force implemented in the single molecule experiments. When the sequence effect is considered, the change of free energy upon each step of the polymerase translocation ($\Delta G = G_{post-trans} - G_{pre-trans}$) plays a similar role as the implemented force such that: $k_t^+ = k_t^+(0)\exp(-\delta \cdot \Delta G / k_B T)$, $k_t^- = k_t^-(0)\exp[(1-\delta)\cdot \Delta G / k_B T]$, in which $k_t^{\pm}(0)$ are the rates at the zero force, and δ is the distance factor ($0 \le \delta \le 1$) for the intermediate state in between the pre- and post-translocated states. Correspondingly, $\sigma_t \equiv \frac{k_t^-}{k_t^+} = \sigma_t(0)\exp(\Delta G / k_B T)$. From Eq (1) one can see that the Michaelis constant $K_M = K[1+\sigma_t]/(1+k_c^+/k_t^+)$ depends on the translocation bias σ_t. Therefore, different sequences (with varying ΔG) can lead to quite different K_M, thus effectively modulate the elongation rate/velocity.

Indeed, the sequence-dependent free energy change ΔG comprises contributions from the melting of the downstream DNA base pair (bp), the re-annealing of the up stream DNA bp, and similar adjustments in the DNA-RNA hybrid (4, 5). Usually, ΔG fluctuates around a value close to zero, with an amplitude ranges from ~1 to 3 k_BT (up to ~ 5 or 6 k_BT at some region) (4). When ΔG is small, $K_M \sim K/(1+k_c^+/k_t^+)$ to $2K/(1+k_c^+/k_t^+)$. A fairly large ΔG leads to a very large $K_M \propto \exp(\Delta G / k_B T)$, therefore, slowing down the elongation significantly at some sequence region. However, since the slowing down happens only at a small portion of the regions along DNA, the overall elongation rate may not be affected much. Below, we numerically examine if the sequence barrier (large ΔG) can slow down the pause-free elongation significantly, as that detected by the single molecule experiment (2).

First we simply assume that there are two types of regions on the DNA: The first type has a fairly small ΔG so that $K_M^I \sim K_M^* \sim 39\mu M$ as measured experimentally (2) (at an assisting force of ~ 6.5 pN with $\sigma_t \sim 4.6$); the second type has ΔG ~6 k_BT such that $K_M^{II} \sim 72 K_M^* \sim 2900\mu M$ (as exp(6)/(1+4.6) ~ 72). For elongation at both types of

regions, the translocation step is assumed much faster than the catalysis step. Suppose that the second type has a population of p and the first type has a population 1-p, so that the combination of both leads to an overall elongation rate as:

$$v = \frac{k_c[NTP]/(1+k_c^+/k_t^+)}{K_M^*+[NTP]} = (1-p)\frac{k_c^+[NTP]}{K_M^I+[NTP]} + p\frac{k_c^+[NTP]}{K_M^{II}+[NTP]} \qquad \text{(S3)}$$

Note that on the LHS of the equation above, the translocation (or force-dependent step) is NOT assumed fast as in (2) ; on the RHS, however, the translocation is assumed fast as mentioned above. At [NTP] ~ 1000 μM, the experimental fitting gives the catalytic rate at ~35 s^{-1} and the slow 'translocation' rate at ~ 112 s^{-1} , so the above equation is numerically written as

$$\frac{1000/(1+35/112)}{39+1000} = (1-p)\frac{1000}{39+1000} + p\frac{1000}{2900+1000} \qquad \text{(S3a)}$$

or even simplified as

$$\frac{1}{1+35/112} \sim (1-p)+p/4 \qquad \text{(S3b)}$$

This way, one estimates p ~ 30%. That is to say, if the large sequence barrier plays a role of significantly slowing down the translocation to a measured forward rate ~ 112 s^{-1} , the large barrier needs to happen at a chance of ~30%. For regular sequences, this population value is much lower (e.g. <5%). For example, if one takes p ~ 2%, and asumming the catalytic rate ~ 35 s^{-1}, the 'translocation' rate k_t^+ is effectively measured at ~ 2300 s^{-1} (on the LHS of Eq S3), much faster than the measured 112 s^{-1}. Hence, regular sequence barrier does not seem to cause an impact as that being detected experimentally (2).

Besides, we performed kinetic Monte-Carlo (KMC) simulation of the elongation, following the three-state BR scheme. The sequence effect is incorporated in the

simulation as one takes ΔG as a normally distributed random variable ~ N(0, σ). Below we show the diagram of the pause-free elongation rate/velocity vs. the sequence variation measure σ. One can see that for realistic values of σ (1~3 k_BT), the pause-free velocities (at [NTP]=1000 μM) remain high (k_c^+~35 s^{-1}). Only when σ reaches ~ 6 k_BT (in case of k_t^+ = 50000 s^{-1}) or 5 k_BT (in case of k_t^+ = 5000 s^{-1}), such that the overall elongation rate reaches to $k_c/(1+ k_c / k_t)$~ 27 s^{-1} at [NTP]=1000 μM (2).

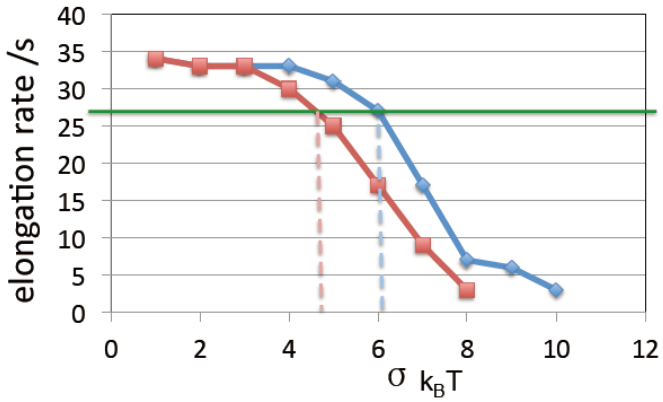


Fig S3. The elongation rate (at [NTP]=1000 μM) vs the sequence stability variation. The data were obtained from the KMC simulation of the three-state Brownian ratchet model of RNAP II elongation (blue, at k_t^+ = 50000 s^{-1} and red, at k_t^+ = 5000 s^{-1}). The sequence-dependent free energy change per translocation step is set as a normally distributed random variable: ΔG ~ N (0, σ). σ is used to describe the sequence stability variation. The green line sets k_{max} ~ 27 s^{-1} as that was measured at this condition (2); the dashed red/blue line thus indicates which value of σ gives the experimental measurement if the slowing down comes from the sequence effect.

The probability that ΔG > σ is ~16% in the normal distribution. Combining the above analyses, it seems that only when higher barriers (5~ 6 k_BT or over) count over ~ 10-30%, can the elongation slowing down become significant as in (2).

IV Estimating lower bound values of the intrinsic NTP dissociation constant K_d for E. coli RNAP

	ATP	UTP	GTP	CTP
k_{max} (s^{-1})	50 ± 6	18 ± 1	36 ± 5	33 ± 6
K^* (μM)	38 ± 7	24 ± 4	62 ± 18	7 ± 4

Table S3 Experimentally measured rate parameters from (6). We use K^* (instead of K_d in the original reference) to avoid confusion of the notations.

Under the BR mechanism, either the 3-state or 5-state model can give a pause-free elongation rate in the form as:

$$v = \frac{k_s / (1 + k_s / k_f)[NTP]}{K[1+\sigma(F)] / (1 + k_s / k_f) + [NTP]} \qquad \text{(S4)}$$

where k_s is the force-independent slow rate (forward rate), and k_f is the force-dependent slow rate (forward rate). In (6), the force-dependent step (or translocation) was assumed very fast, and the above elongation rate is written as:

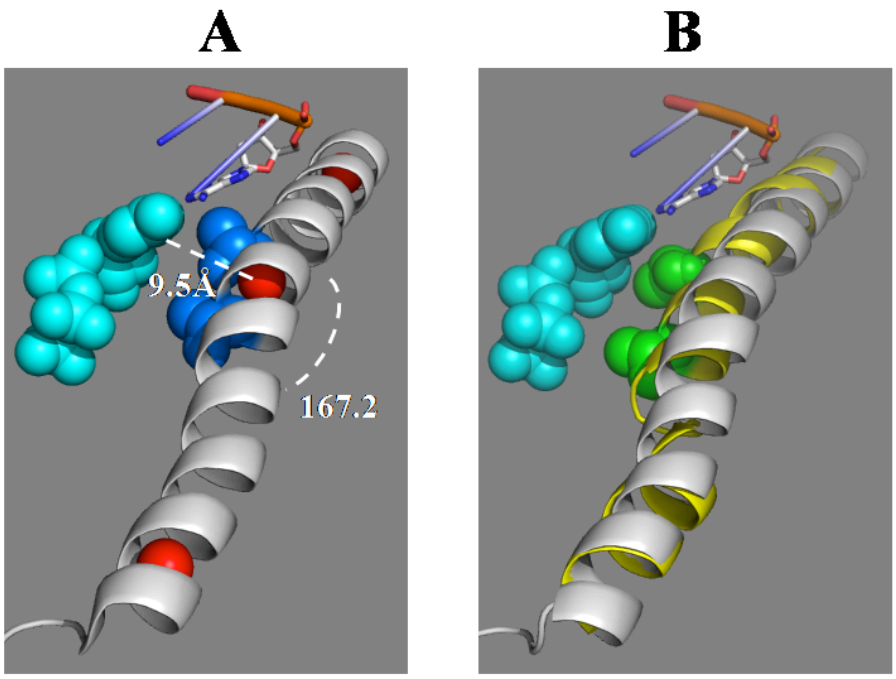
$$v = \frac{k_{\max}[NTP]}{K^*[1+\sigma(F)] + [NTP]} \qquad \text{(S5)}$$

with $k_s = k_{max}$ if $k_s << k_f$. Hence, the k_{max} and K^* measured in (6) correspond to $k_s/(1+k_s/k_f)$ and $K/(1+k_s/k_f)$ in Eq (S4). From Eq (1) and (4) in main: $K=K_d + k_s / k^0_b$ or from Eq (5) in main: $K=K_d\, \sigma_{TLc}/(1+\sigma_{TLc}) + k_s/ k^0_b$. Below we still use k_s/k_f ~0.3 as that in the the generic BR model (2).

From the data of CTP in **Table S3**, one obtains $k_s/k^0_b < K=7^*(1+ k_s/k_f)$~ 9 μM, k_s~33*（$1+k_s/k_f$ ）~43 s^{-1}, so that $k^0_b >$ ~5 $\mu M^{-1}s^{-1}$. Since the NTP binding is dominated by a diffusion process, the binding rate constant k^0_b is regarded the same for all species of NTP. If we use $K=K_d + k_s/k^0_b$ as that in Eq (1) and (4), then K_d for the rest of NTP

species can be estimated as K_d=K- k_s/k^0_b = K*(1+ k_s/k_f) - k_{max}(1+ k_s/k_f)/ k^0_b > 36 μM for ATP (as k^0_b > 5 $\mu M^{-1}s^{-1}$), K_d > 25 μM for UTP, and K_d > 72 μM for GTP. It seems that these values are at odds with K_d < 10 μM inferred from **model A** in main for RNAP II (either Eq (1) or (4) works). Hence, if *E.coli* RNAP and Pol II have similar values of K_d and elongation mechanisms, then **model B** seems consistent with both sets of the measurements from (2) and (6).

VI. Blockage of the NTP binding A-site during translocation of RNAP II from MD simulation



FigS4. (A) Crystal structure of the UTP-bound Pol II complex (PDB id: 1R9S) that was used as the reference for the analysis. UTP (in cyan), BH (in gray), BH residue T827 and T831 (in blue) are shown. The reference angle is defined as the center of mass (in red) of three helical turns (in orange) from the BH, which is determined to be 167°. And the reference distance is defined as the distance between the base group of the UTP and the middle helical turn defined above, which is determined to be ~9.5 Å and indicated with an orange dashed line. (B) One snapshot from the MD simulations (in yellow) superimposed to the crystal structure showing that the BH bending may affect the NTP binding site, residue T827 and T831 are shown in green.

State	Equilibrium Population	Bending Population	Blockage Population
S1	49.1%	79.2%	70.4%
S2	17.3%	33.5%	30.5%
S3	23.7%	63.5%	62.5%
S4	9.9%%	20.6%	6.1%

Table S4 Statistics related to bending configurations of BH during translocation, sampled from MD simulation (3). The 'Equilibrium Population' provides equilibrium probability distributions of the pre-translocated (S1), intermediate (S2 and S3), and post-translocated state (S4). The 'Bending Population' gives the percentile of the BH bending (the above-defined angle is smaller than 167°) vs the straightening within each equilibrium configuration, while the 'Blockage Population' gives the chance that the bent configuration would also block the NTP binding A-site (the above-defined distance is shorter than 9Å).

References

1. Larson, M. H., J. Zhou, C. D. Kaplan, M. Palangat, R. D. Kornberg, R. Landick, and S. M. Block. 2012. Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. Proceedings of the National Academy of Sciences 109:6555-6560.
2. Dangkulwanich, M., T. Ishibashi, S. Liu, M. L. Kireeva, L. Lubkowska, M. Kashlev, and C. Bustamante. 2013. Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. eLIFE 2:e00971.
3. Silva, D.-A., D. R. Weiss, F. Pardo Avila, L.-T. Da, M. Levitt, D. Wang, and X. Huang. 2014. Millisecond dynamics of RNA polymerase II translocation at atomic resolution. Proceedings of the National Academy of Sciences.
4. Bai, L., A. Shundrovsky, and M. D. Wang. 2004. Sequence-dependent Kinetic Model for Transcription Elongation by RNA Polymerase. Journal of Molecular Biology 344:335-349.
5. Yager, T. D., and P. H. Von Hippel. 1991. A thermodynamic analysis of RNA transcript elongation and termination in Escherichia coli. Biochemistry 30:1097-1118.
6. Bai, L., R. M. Fulbright, and M. D. Wang. 2007. Mechanochemical Kinetics of Transcription Elongation. Physical Review Letters 98:068103.